Can Top-of-Atmosphere Radiation Measurements Constrain Climate Predictions? Part II: Climate Sensitivity

Can Top of Atmosphere Radiation Measurements Constrain Climate Predictions? Part 2: Climate Sensitivity.

Simon F. B. Tett *

School of GeoSciences, University of Edinburgh

Daniel J. Rowlands

Atmospheric, Oceanic & Planetary Physics, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK

Michael J. Mineter

School of GeoSciences, University of Edinburgh

Coralia Cartis

School of Mathematics, University of Edinburgh, The King's Buildings, West Mains Road, Edinburgh EH9 3JZ, UK

ABSTRACT

A large number of perturbed-physics simulations of the HadAM3 atmospheric model were compared with the CERES (Clouds and Earth's Radiant Energy System) estimates of Outgoing Longwave Radiation (OLR) and Reflected Shortwave Radiation (RSR) as well as OLR and RSR from the earlier ERBE (Earth Radiation Budget Experiment) estimates. The model configurations were produced from several independent optimisation experiments in which four parameters were adjusted. Model-observation uncertainty was estimated by combining uncertainty arising from: satellite measurements, observational radiation imbalance, total solar irradiance, radiative forcing, natural aerosol, internal climate variability, Sea Surface Temperature and that arising from parameters we did not vary. Using an emulator built from 14,001 "slab" model evaluations carried out using the climate*prediction*.net ensemble the climate sensitivity for each configuration was estimated. Combining different prior probabilities for model configurations with the likelihood for each configuration, and taking account of uncertainty in the emulated climate sensitivity gives, for the HadAM3 model, a 2.5-97.5% range for climate sensitivity of 2.7-4.2 K if the CERES observations are correct. If the ERBE observations are correct then they suggest a larger range, for HadAM3, of 2.8-5.6 K. Amplifying the CERES observational covariance estimate by a factor of 20 brings CERES and ERBE estimates into agreement. In this case the climate sensitivity range is 2.7-5.4 K. Our results rule out, at the 2.5 % level, for HadAM3 and several different prior assumptions climate sensitivies greater than 5.6 K.

1. Introduction

Considerable uncertainty exists about how sensitive the climate is to changes in CO2. This is often summarised as "equilibrium climate sensitivity" (S): the equilibrium global-average temperature change in response to a doubling of CO2. The most recent IPCC assessment reported that S was *likely* (more than 66% chance) to be in the range 2.0 to 4.5K and values greater than 4.5K could not be ruled out(Meehl et al. 2007). This uncertainty largely arises from uncertainty in modelling atmospheric processes such as cloud formation and convection as well as changes in snow and ice which act to modify the "greenhouse" effect and albedo of the planet. Uncertainties in climate sensitivity combined with uncertainty in the rate at which the oceans take up head lead to uncertainty in the response of the climate system to changes in greenhouse gases.

Knutti and Hegerl (2008) review various approaches to provide probabilistic estimates of equilibrium climate sensitivity. In broad terms these can be classified into two different approaches. One in which observed change or variability over the last few decades to las millennium have been compared to models of varying complexity (e.g. Hegerl et al. (2006); Kettleborough et al. (2007); Olson et al. (2012)). A second approach has been to compare observed and simulated climatologies(e.g. Sexton and Murphy (2011); Stainforth et al. (2005); Murphy et al. (2004); Sanderson (2011)). Such approaches are often multi-variate and assess observational error by using multiple observational datasets. Huybers (2010) explored climate sensitivities in the CMIP3 archive and found some evidence

that models had been tuned as there was compensation between feedbacks arising from different processes. Lemoine (2010) carried out a similar analysis but considered common biases between models and found considerable sensitivity to assumptions about these biases. Huber et al. (2011) analysing results from the CMIP3 archive and comparing with radiation measurements found ranges of climate sensitivity from 2.9-4.5K though did not attempt a probabilistic estimate.

In perturbed physics ensembles(Murphy et al. 2004) key parameters in a climate model are varied within their uncertainty ranges leading to the possibility of climate sensitivities(Stainforth et al. 2005) much larger than 6K. Recently, Rowlands et al. (2012) reported on an observationally constrained rates of warming to 2050 by comparing a very large ensemble of perturbed-physics HadCML simulations with observations for the period 1960-2010 period. They concluded that global mean warming in the 2050s, relative to 1961-1990, is likely in the range 1.4-3K

Jackson et al. (2008) using a improved Markov-chain monte-carlo algorithm generated a range of perturbed parameter versions of the CAM3.1 model. They used a range of observations, largely based on reanalyses, to constrain the plausible parameter choices and found that model configurations with the smallest systematic errors had a climate sensitivity within 0.5K of 3K. Järvinen et al. (2010) also applied a variant of the Markov-chain monte-carlo algorithm to estimate parameters though did not draw inferences on climate sensitivity.

In part 1 of this paper we reported on successful attempts to automatically tune the 3rd Hadley Atmospheric Model (HadAM3; Pope et al. (2000)) at its N48 ($3.75^\circ \times 2.5^\circ$) resolution to the Loeb et al. (2009) 2000-2005 observations of Top of Atmospheric (TOA) radiation values. In this paper we use results from those simulations to draw observationally constrained inferences about climate sensitivity. In an atmospheric model with fixed Sea Surface Temperatures (SST) we hypothesise that there is a relationship between TOA radiation and climate feedbacks. We put forward this hypothesis as the processes that cause climate feedbacks also modify the outgoing radiation budget in a fixed SST simulation. For example water vapour is transported into the upper troposphere via convection and there it reduces OLR in the fixed SST experiment. Changes in water vapour in response to atmospheric temperature changes through convective transport are also one possible climate feedback. Tropospheric water vapour also produces clouds which effect the radiation balance of a model and changes in cloud in response to climate change are a significant climate feedback.

The aims of this paper are to:

i. Explore if there is any relationship between simulated outgoing radiation and climate sensitivity for HadAM3.

ii. Explore if this relationship is one-to-one or one-to-many.

iii. Produce an uncertainty estimate for model-observational discrepancy.

iv. Use this estimate to produce an uncertainty estimate for equilibrium climate sensitivity based on the use of the HadAM3 model.

The rest of the paper is structured as follows. In the following methods section we first describe our modelling strategy, how we estimate climate sensitivity using an emulator based on data from the climate*prediction*.net ensemble (Sanderson et al. 2008), then present our comprehensive analysis of uncertainties on model-observational discrepancies before, finally, describing how we compute cumulative density functions for climate sensitivity. Having described our methods we then present our results before concluding with an extended discussion.

2. Methods

In this section we first briefly describe our modelling strategy, and how we estimated uncertainty in model-observational discrepancy. We then describe how we estimated climate sensitivity for any parameter combination using an "emulator" (e.g. Rougier (2007)) and how we used those to compute cumulative distribution functions for climate sensitivity.

a. Modelling

The default configuration of HadAM3 has been evaluated(Pope et al. 2000) and has an estimated climate sensitivity of 3.3K to doubling CO_2 (Williams et al. 2001; Randall et al. 2007). This estimate disagrees with the estimate of 3.7K of Gregory and Webb (2008) because their calculation was done by halving the response to 4xCO_2 while Williams et al. (2001) doubled CO_2. Our climate sensitivity estimates (see later) are based on the response to doubling CO_2. For our atmospheric only simulations we modified the default model to include a package of natural and anthropogenic forcings(Tett et al. 2007), included a recent estimate of total solar irradiance(Kopp and Lean 2011) and corrected a bug in the Rayleigh scattering shortwave coefficients.

We "tuned" the N48 ($3.75^\circ \times 2.5^\circ$) version of HadAM3 by modifying four parameters (`entcoef`, `vf1`, `ct` and `rhcrit`) which previous work(Knight et al. 2007) had shown were important for climate sensitivity. The parameters were modified using an optimisation algorithm which aimed to produce models with specified global-mean outgoing longwave radiation (OLR) and reflected solar radiation (RSR).

We carried out several optimisation experiments. For each experiment we started from 16 different extreme com-

binations of the four parameters and optimised each starting parameter choice to the same target values of OLR and RSR. We optimised to six targets in all. In all, we carried out about 2500 simulations of HadAM3 which have a broad range of OLR and RSR values though with highest density around the observed values. The parameter values for those configurations close to the observed targets had a broad range of values.

In part 1 we showed that there was a compensation in clear sky OLR between upper tropospheric temperature and water. Consequently most of the changes seen in the ensemble were driven by changes in cloud. This is consistent with literature on reasons for uncertainty in climate sensitivity (e.g. Webb et al. (2006); Randall et al. (2007)). More details can be found in part 1 of this paper.

b. Observational-Model Discrepancy

In this paper we concentrate on comparison with the recent CERES observations of OLR and RSR (Loeb et al. 2009) but also consider the older ERBE values (Fasullo and Trenberth 2008). Any comparison between observations and models requires a quantitative estimate of observational-model discrepancy. We make this estimate by considering several sources of uncertainty and combining them to produce a total uncertainty. Our focus is on observational uncertainties and modelling uncertainties which affect outgoing radiation. In order to convert some sources of uncertainty to uncertainties in OLR and RSR we make use of some HadAM3 simulations. Unless stated otherwise these make use of the standard configuration of HadAM3 and do not consider structural uncertainty. As we are considering large scale processes we assume that all uncertainties are Gaussian and make estimates of their values for the 2001-2005 period. We consider the following sources of uncertainty quantifying them as plus/minus one standard deviation:

Satellite Measurement From Table 2 of Loeb et al. (2009) we sum the individual components of the bias uncertainty to give a total observational uncertainty of 1.4 and 1.0 Wm^{-2} for OLR and RSR respectively. We assume these are independent of one another.

Observational radiative imbalance The TOA radiation dataset we use was adjusted to have the same radiation balance(Loeb et al. 2009) as ocean observations(Willis et al. 2004) for the upper 750 m (0.86 ± 0.12 Wm^{-2}). Lyman et al. (2010) estimated the upper ocean has warmed by 0.55 to 0.73 Wm^{-2} . We use $0.75 \pm 0.25 \mathrm{Wm}^{-2}$ to include both estimates.

Uncertainty in incoming radiation There is a small uncertainty in the incoming TSI from which we derive the balance requirement. Solar minimum TSI(Kopp and Lean 2011) has been estimated at 1360.8 ± 0.5 Wm^{-2} significantly different from older estimates(Willson and Hudson 1991). For the period 2001-2005 this gives a incoming top of atmosphere (TOA) radiation of 1362.2 ± 0.5 Wm^{-2} arising from the elliptical nature of the Earth's orbit and slightly higher TSI when the sun is active.

Internal Climate Variability From an ensemble of 19 standard configurations of HadAM3 we estimated the covariance of OLR and RSR. The standard deviations are about 0.1 Wm^{-2} for both. This is a negligible source of uncertainty and reflects our use of atmospheric models rather than coupled atmosphere/ocean models where variability in outgoing radiation is much larger (for example Tett et al. (2007)).

Forcing uncertainty Our simulations are all driven with a package of radiative forcings which are uncertain(Forster et al. 2007). Forcing is the change in downward radiative flux at the tropopause after stratospheric adjustment(Tett et al. 2002) when the stratosphere is in equilibrium. As our simulations are driven with observed sea surface temperatures feedbacks between forcing and atmospheric state will be reduced and so we'd expect a change in forcing to produce a similar change in OLR and RSR as the LW and SW forcing respectively.

To test if forcing and outgoing radiation were similar we carried out a set of three simulations using the standard configuration of HadAM3: **NOSUL** in which we removed the direct and indirect effect of sulfate aerosols; **Natural** in which all anthropogenic forcings were removed; **NATGHG** in which all anthropogenic forcings except well mixed greenhouse gases (GHG) are removed. We then compared these values for 2001-2005 with the forcing calculations of Tett et al. (2007) for 2000. We expect the forcing from 2000-2005 to be similar and errors in the forcing estimate are small. The changes in OLR and RSR for **Natural** are broadly consistent with the forcing estimates (Table 1) with a slightly larger OLR than the forcing would suggest. However, considering the changes in the **NOSUL** and **NATGHG** experiments suggests this may be arising from some compensation between the effect of aerosols and GHG on OLR and RSR. Some of this may be due to internal variability in the model simulations.

The dominant SW forcing arises from aerosol forcing and so being cautious we estimate, by comparison between **NOSUL** and the reference simulations, that this reduces RSR by $1.2\mathrm{Wm}^{-2}$ and increases OLR by 0.2 Wm^{-2} . We compute the covariance matrix for a 1 Wm^{-2} uncertainty in SW forcing by computing the outer matrix product) of this vector with itself (the

matrix multiplication of a vector by its transpose so $C_{ij} = v_i v_j$) scaled by 1/0.9 (the SW forcing). We assume that LW forcing largely arises from changes in well mixed greenhouse gases and compute their impact on RSR and OLR as 0.7 and 2 Wm^{-2} respectively from the difference between **NATGHG** and **Natural** simulations. We then scale these values by 1/2.11 (the LW forcing) and compute the covariance from the outer matrix product to obtain a covariance for a change in LW forcing of 1 Wm^{-2} .

To obtain radiative forcing uncertainties we use the existing uncertainty estimates (Table TS.5 of Solomon et al. (2007)), assume uncertainties are independent and round to one significant figure to give 1σ uncertainties of 1 and 0.20 Wm^{-2} for RSR and OLR. These are then used to scale the covariance matrices computed above. The dominant contribution to RSR are from uncertainty in the direct effect and cloud albedo effects of aerosol while for OLR the dominant uncertainties arise for CO_2 and ozone forcing. Our estimates of forcing uncertainty, to some extent, includes structural uncertainty as we use the range of forcing values from Forster et al. (2007) though modified through use of HadAM3 to obtain TOA RSR and OLR.

Natural Aerosols There are many natural aerosols in the climate system which largely effect RSR with a minimal impact on OLR(Carslaw et al. 2010). For the current climate natural aerosol feedbacks on the radiation budget are small so we use uncertainties in natural aerosols directly. Key components are organic aerosol, aerosol from the impact of fire, and dust whose effect on RSR has been estimated at -0.03 to -1.1, -0.05 to 0.2 and -0.7 to 0.5 Wm^{-2} respectively(Carslaw et al. 2010). These would combine, assuming independence and that the estimates are 5-95% Gaussians, to a total 1σ uncertainty of about 0.6 Wm^{-2} . We also used three simulations from Penner et al. (2006) and after correcting all contemporary simulations to the same RSR found the range in pre-industrial RSR was 1 Wm^{-2} . We used this as a 1σ estimate for natural aerosols as it was greater than the Carslaw et al. (2010) estimate. As with forcing uncertainty this uncertainty range incorporates structural uncertainty in the impact of aerosols on outgoing radiation.

SST Uncertainty On the 5-year timescales we are considering the major source of SST uncertainty is the climatology rather than uncertainty in the individual years. Two Hadley SST datasets have climatological differences of less than 0.2K(Rayner et al. 2006) over most of the world ocean. Therefore, we assume the 1σ in SST is 0.2K. We estimated its impact on RSR and OLR by, everywhere, increasing the SST values by 0.5K and forcing default HadAM3 with it. We found a change in RSR and OLR of -0.4 and 1.2 Wm^{-2} and scaled the response by 2/5 to give the covariance for SST uncertainty of 0.2 K. We also carried out a simulation using the SST dataset of Reynolds et al. (2002) and found this had a small impact on RSR and OLR (about 0.1 Wm^{-2}).

Other parameters Our results are based on modifying four HadAM3 parameters which most affect S. Other parameters have less effect on S but could affect the outgoing radiation. For example, one of our configurations could be inconsistent with the observations but if we modify other parameters it may be. We treated this as another source of uncertainty. To estimate its covariance we found the 13 distinct parameter combinations from the 14,001 climate*prediction*.net cases that had default values for `entcoef`, `vf1`, `ct` & `rhcrit` and a climate sensitivity between 3.2 and 3.4. We then ran them to compute the RSR and OLR (Fig. 1) and computed a covariance matrix from the 13 cases. The parameters (Knight et al. 2007) that varied were `cw` (precipitation threshold), `ice_size` (size of ice particles),`dtice` (ice albedo variation), `alpham` (ice albedo at melting point) and `ice` (non-spherical ice). The largest changes in RSR and OLR arose from using non-spherical ice. However, these parameter changes didn't greatly change the total outgoing radiation.

Loeb et al. (2009) computed estimates for average RSR and OLR by adjusting the measured RSR and OLR values within their estimated bias uncertainties until they were consistent with the ocean heat content estimates of net imbalance. As we are using slightly different estimates of ocean heat content uncertainty and require a covariance estimate we computed observational uncertainty by combining distributions for the individual outgoing SW and LW radiation bias uncertainties with a distribution for the total radiation. Total outgoing radiation is taken to be the total incoming - the expected imbalance ($(1362.2/4 - 0.75) \pm 0.3$ $\mathrm{W/m}^2$). We extended this to include uncertainty in the orthogonal and independent component – the difference between RSR and OLR. In the absence of other information we assume the distribution for the difference is a normal distribution with mean -40% (corresponding to an albedo of 0.3) and standard deviation 10% of the incoming radiation. These are large enough that other uncertainties dominate. We combine this covariance with the satellite bias covariance by first computing the precision matrix (inverse of covariance matrix), linearly transforming the precision matrix to give the individual RSR and OLR components and combining with the precision matrix for satellite bias

uncertainty through the formula:

$$\begin{aligned}\Lambda_c =& \Lambda_L + \Lambda_b \\ \mu_{\mathbf{c}} =& \Lambda_c^{-1}(\Lambda_L \mu_{\mathbf{L}} + \Lambda_b \mu_{\mathbf{b}})\end{aligned}$$

where Λ is the precision matrix and μ is the mean value. Subscripted $_c$, $_b$ and $_L$ are the combined values, near-radiative balance covariance (defined above) and those from Loeb et al. (2009) respectively.

Our analysis gives $\mu_{\mathbf{c}} = (99.7, 240.0)\mathrm{W/m^2}$ for the RSR and OLR slightly different from $(99.5, 239.6)\mathrm{W/m^2}$ of Loeb et al. (2009). We estimate the covariance matrix of this combined observational error (ocean heat content and satellite bias error) to be:

$$\begin{pmatrix} 0.74 & -0.64 \\ -0.64 & 0.82 \end{pmatrix}.$$

Other sources of uncertainty are assumed independent of each other and added to this covariance matrix. The different sources of uncertainty vary in their magnitude though total modelling (all sources of uncertainty considered apart from satellite bias and ocean heat content) uncertainty is much larger than observational uncertainty. The most important contributors to total uncertainty come from forcing and parameter uncertainty (Fig 1) with a total covariance matrix estimated to be:

$$\begin{pmatrix} 7.6 & -4.5 \\ -4.5 & 4.3 \end{pmatrix}.$$

c. Estimating Climate Sensitivity

Running full simulations to calculate the equilibrium climate sensitivity for each parameter combination was not possible given computational constraints. We have therefore used a statistical model to estimate the climate sensitivities for each of the candidate parameter combinations produced in our optimisation algorithm. This approach is becoming widely adopted in the field, whereby a statistical model can be trained on past evaluations of a climate model with perturbed physics and then used to predict various output quantities for new parameter combinations (Sanderson et al. 2008; Rougier et al. 2009; Sanso and Forest 2009), with the term *emulator* being adopted. The recent UKCP09 (UK Climate Projections – see `http://ukclimateprojections.defra.gov.uk/`) climate projections relied heavily on the use of statistical emulation (Murphy et al. 2007). These algorithms can simply be thought of as non-linear regressions of the climate model parameters onto output quantities of interest.

In this study we have used the randomForest technique (Breiman 2001) to build our statistical emulator, based on a 14,001 member perturbed physics ensemble generated from climate*prediction*.net. All simulations were from HadSM3, which consists of the same atmospheric model coupled to a slab thermodynamic ocean. Climate sensitivities were estimated using existing methodology (Stainforth et al. 2005). Subsequent simulations have recently been performed to vary parameters continuously rather than the original grid design improving the ability of our emulator to learn about how climate sensitivity changes as we vary the model parameters. In all 10 parameters were varied independently in the climate*prediction*.net ensemble(Sanderson et al. 2008), of which 4 of the most influential are considered here.

The randomForest technique is a machine learning algorithm which has been shown to be very powerful in capturing non-linear dependencies in a wide-variety of problems(Breiman 2001). The algorithm constructs an ensemble of regression trees each built with a bootstrap sample of the original training data, with randomised splitting at each node. The aggregation of a number of classifiers together, known as Bagging (Bootstrap Aggregating), vastly improves the performance of the algorithm and avoids over fitting. The algorithm requires only 3 parameters in the setup, namely the number of regression trees, terminal node size and number of parameters to split the data over at each stage in the tree construction. Sensitivity studies (not shown) indicate that in this case the results of randomForest estimates of climate sensitivity are not significantly changed by varying these parameters.

Figure 2a) shows the performance of the predictions from the randomForest comparing simulated values in HadSM3 to predicted values from the randomForest algorithm. All of these predictions are *out-of-sample*, meaning that predictions were made for models not included in the process of fitting the randomForest. This was achieved through a 10-fold cross validation as follows,

- Randomly split the 14,001 member ensemble up into 10 segments.
- Remove all models in segment 1 from the ensemble, and fit the randomForest on the remaining 90%. Once fitted make a prediction for the models that have been left out.
- Repeat for segment 2 and so on until every segment has been left out.

The result is a set of 14,001 predictions for climate sensitivity, which can be compared to the actual simulated values from HadSM3. We do not expect the randomForest to perfectly fit the simulations since the climate sensitivity values we fit to are contaminated by noise due to internal variability and uncertainties induced by the exponential fit used (Stainforth et al. 2005). Overall we find that the randomForest predictions explain over 95% of the variability in the simulated climate sensitivity values.

We account for uncertainty in our climate sensitivity estimates by using the error statistics generated from the out-of-sample predictions. Specifically we calculate the root mean square error (RMSE) in bins of simulated climate sensitivity (Figure 2b). The bins are chosen to span the 5-95% range of the simulated climate sensitivity distribution in deciles. This RMSE is approximately 0.2K at climate sensitivities of 2-3K and rising to 0.6K for climate sensitivities above 6K, which we use as a varying 1σ error in our analysis. Integrated over all climate sensitivities the RMSE is approximately 0.3K (Figure 2c). These values are in line with estimates of the uncertainty from initial condition ensembles (Stainforth et al. 2005). Our error estimates are approximately 50% smaller than a similar study who found a 1σ error of approximately 0.45K (Rougier et al. 2009). We attribute this to our ensemble being approximately 50 times larger and exploring a lower dimensional parameter space leading to a much denser sampling of points.

Other methods exist to estimate uncertainty in predictions from the randomForest technique (Meinshausen 2006). We use out-of-sample error statistics as it is a standard statistical technique often used in ensemble forecasting (Roulston and Smith 2003), that simplifies and adds transparency to our analysis.

The impact of sampling uncertainty, that is the uncertainty in the predicted climate sensitivity arising from the specific ensemble members used to fit the randomForest, is very small relative to the prediction error (Fig. 2(b)) and so is ignored in our uncertainty analysis.

d. Computing Probability Density Functions

In part 1 of this paper we showed that an optimisation method can be used to adjust model parameters so as to produce simulated global averages of reflected solar radiation (RSR) and outgoing longwave radiation values (OLR) that are close to target values. We carried out about 2500 simulations in all and they have a broad range of OLR and RSR values. We would like to use results from those simulations to make probabilistic statements about the climate sensitivity of HadAM3. We chose to focus on climate sensitivity as it is a key summary parameter for future climate change. However, our method could be applied to any future prediction of a climate modelling system.

One issue we face is that the configurations we use are generated by an optimisation algorithm. The algorithm has the advantage of generating configurations that are close to the target values but at the cost of making configurations dependant on each other. From the approximately 2500 cases we had 78 configurations with an root-mean-square-error of less than 1 Wm^{-2} to either the CERES or EBRE observations. The individual parameter values, from this sub-set, cover a broad range of values suggesting we are sampling the distribution well. However, as discussed in part 1 there are correlations between the different parameter values.

We label each one of N model configurations $\mathcal{M}_i$ with a simulated RSR and OLR of $\mathbf{r_i}$. Each model configuration has an associated climate sensitivity (S_i) which we compute using the emulator described in Section c. Using Bayes theorem we can write the probability of a model configuration given observations $P(\mathcal{M}_i|O)$ as:

$$P(\mathcal{M}_i|O) \propto P(O|\mathcal{M}_i)P(\mathcal{M}_i) \tag{1}$$

The constant of proportionality can be computed by requiring that the probabilities sum to 1. $P(O|\mathcal{M}_i)$ is the likelihood (L_i) of $\mathcal{M}_i$ and we now describe how this is computed. The density of model configurations near target values is largest while those far away from the target have low sampling densities. The probability **density** for a multi-variate Gaussian distribution with mean μ and Covariance $\mathbf{C}$ is:

$$\rho(\mathbf{r}) = \frac{1}{(2\pi)\sqrt{\det \mathbf{C}}} \exp\left(-\frac{1}{2}(\mathbf{r}-\mu)\mathbf{C}^{-1}(\mathbf{r}-\mu)^T\right) \tag{2}$$

We assume that the likelihood of each $\mathcal{M}_i$ varies smoothly in in a small patch (Ω_i) around $\mathbf{r_i}$ and we can compute likelihoods using:

$$L_i = \int_{\Omega_i} \rho(\mathbf{r}) d\mathbf{r} \tag{3}$$

For small enough Ω_i then $\rho(\mathbf{r})$ is approximately constant ($= \rho(\mathbf{r_i})$) giving:

$$L_i = \rho(\mathbf{r_i}) A_i \tag{4}$$

where $A_i = \int_{\Omega_i} d\mathbf{r}$. We compute A_i from the area of the Voronoi polygon (Aurenhammer 1991) for $\mathbf{r_i}$. The Voronoi polygon is the polygon that surrounds the region which is closer to $\mathbf{r_i}$ than any other $\mathbf{r_j}$. These ideas can be generalised to higher dimensions.

Within the 99% consistency region a small number of the Voronoi polygons have large areas (Fig. 3). These polygons include the boundaries of the region generated by our optimisation process and so their area depends on the arbitrary choice of boundary. Over such large polygons the area will no longer be sufficiently small that we can approximate Eqn. (3) by Eqn. (4). So we cap the polygon area at π corresponding to a circle with unit radius giving zero likelihood to regions far from points we generated. We explored reducing this cap to $\pi/4$ and found it made little difference to our results. Our sampling across the 99% consistency region is sufficiently dense, except near some of the boundaries (Fig. 3), that our assumptions appear reasonable.

Our climate sensitivity values are based on integrations of a few decades of a "slab" climate model and use of an emulator and so are uncertain. This uncertainty depends

on S itself (Section c and Fig. 2(b)). We assume, to be conservative, that this emulator uncertainty is coherent across all $\mathcal{M}_i$ and thus make emulator uncertainty a significant contribution to total uncertainty. To include this uncertainty we generate 10 random realisations of S assuming the uncertainty in it is Gaussian. For each realisation we generate a **single** realisation (ϵ) from a Gaussian distribution. Based on the uncertainties shown in Fig. 2 we add 0.2ϵ when $S_i < 3.5$ to the emulated climate sensitivity. For values of S larger than this we add: 0.3ϵ when $3.5 <= S_i < 4.5$; 0.4ϵ when $4.5 <= S_i < 6$; 0.5ϵ when $6 <= S_i$. We then re-normalise the existing likelihoods, priors and posteriors and computed cumulative distributions for S from these distributions.

3. Results

We now apply the uncertainty estimates and approaches described above to computed cumulative distribution functions for Climate Sensitivity. Before doing that we revisit the groups we used in part 1 of this paper.

In part 1 we split the configurations close to the CERES and ERBE observations into two groups on the basis of their land temperatures. Both groups had simulated OLR and RSR close to the target values but one group was warmer over land and drier in the tropics (termed CERES or ERBE warm group) than the Standard configuration. The other cluster (termed CERES or ERBE cold group) had a surface climatology close to the default configuration. The CERES cold group has a mean S of 3.2 K slightly smaller than the standard HadAM3 sensitivity(Williams et al. 2001). The CERES warm group has a slightly higher sensitivity of 3.6 K. The ERBE clusters have a broader range of sensitivities of 3.3 K(cold group) and 4 K(warm group). This suggests that it is possible to get different climate sensitivities for similar values of OLR and RSR and so S is not a single valued function of OLR and RSR.

Using all our simulations and the emulated climate sensitivity we can explore the dependency of S on OLR and RSR. This shows that high values of S occur for low values of RSR while smallest values occur at high values of both OLR and RSR (Fig. 4). Zooming in closer to the observations we can see a rich structure with islands of high sensitivity surrounded by regions of lower sensitivity (Fig. 5). We can also see quite a dense sampling of model configurations in the region where configurations would be consistent at the 95% level.

Our original simulations had a S range of 2.5 to 10.2K. Only a subset of these are consistent with observations, at the 95% level, with climate sensitives ranging from 3.0 to 4.1K for the CERES observations and 3.0-5.2 K for the ERBE observations. To build a coupled ocean/atmosphere model we would require that the atmospheric model be in near radiative balance which we interpret as within a Wm^{-2} of the observed value. The model configurations that had an imbalance within 1 Wm^{-2} of the observed value had a S range of 3.0 to 4.1 K for CERES and 3.0-5K for ERBE. This suggests that, for HadAM3, it is possible to build coupled models that do not need flux correction but which span a plausible range of climate sensitivities.

Given the model configurations are not uniformly sampled nor randomly generated our approach is to take five different prior distributions and then compute five posterior distributions. If the posterior distributions are similar then the observations are important constraints on the posterior probabilities.

We use equal-probable prior distributions where some property is equally likely within the range of simulated values. The five we consider are:

Uniform All configurations are equally likely. This gives a posterior probability equal to the likelihood.

Radiation All values of OLR and RSR are equally likely.

Parameter All parameter values are equally likely.

S All climate sensitivities are equally likely.

1/S All climate feedback values (1/S) are equally likely.

For equal-probable climate sensitivity we computed the weights from the difference in the climate sensitivities with the boundary values having the same weight as those interior. Climate sensitivities were ordered monotonically prior to computing the weights. Similar computations were done for equal-probable climate feedbacks. For Radiation and Parameter posteriors we constructed the priors from the area/4-volume of the Voronoi polygons/4-polytopes. For radiation weights we capped the area of the polygon at π. For Parameter weights we capped the 4-volume of the 4-polytopes at 1000 times the median 4-volume of the Voronoi polytopes.

The priors we considered lead to a range of different cumulative distribution functions (Fig. 6(a)). When combined with the observations, and our estimated uncertainty, the posterior distributions are all very similar (Fig. 6(a)) with 95% of climate sensitivities between 3 and 4K. Taking account of uncertainty in the emulated climate sensitivity leads to a broader distribution function (Fig. 6(b)) again with little sensitivity to prior assumption. This suggests, that, for our uncertainty estimate and HadAM3, the CERES observations provide a strong constraint on climate sensitivity. Examining the 2.5%, best estimate and 97.5% values of the distribution (Table 2) shows that sensitivity to prior assumption is about 0.1 giving climate sensitivities ranging from 2.7-4.2K for HadAM3 with a best estimate of 3.4K. Climate sensitivities outside this range are inconsistent with the CERES observations.

We then explored how sensitive our results are to different assumptions. These sensitivity experiments are:

ERBE We treated the ERBE values of Fasullo and Trenberth (2008) as we did the CERES results. We first scaled the ERBE RSR value by the Kopp and Lean (2011) TSI values divided by 1365 Wm^{-2} . As modelling uncertainty dominates our total covariance we used the same covariances as before in our analysis.

CERES 2× We scaled the covariance matrix generated from our uncertainty analysis by a factor of two but used the CERES observations.

CERES 20× **Sat.** We scaled the Loeb et al. (2009) RSR and OLR covariance by a factor of 20 and used the CERES observations. This is sufficient (Fig. 5) to make the ERBE and CERES values consistent with one another.

2002 CERES We only used simulated data for 1st December 2001 -30th November 2002 in our observational-model comparison. Internal variability was computed from the 19-member ensemble of HadAM3. The CERES observations and other contributions to total uncertainty were as the base case.

CERES Sample We only used the first one of seven simulations in each of the optimisation iterations (see part 1). The same observations and uncertainties were used. This should increase the independence of the samples.

For each of these sensitivity studies we repeated our earlier analysis. Using the ERBE observations rather then the CERES observations has a large impact (Fig. 7 and Table 2) with much greater sensitivity to prior assumptions, a marginally increased lower bound for climate sensitivity and a much increased upper bound for climate sensitivity. Using the ERBE results we would report a 2.5-97.5 % climate sensitivity range of 2.8-5.6K with best estimates around 4K.

Increasing the covariance and using the CERES observations (CERES 2×), not unexpectedly, increases the range of plausible climate sensitivities with a small impact at the lower end but increases the upper end to 5K. It also increases the sensitivity of our results to prior assumptions. Increasing the CERES measurement uncertainties (CERES 20× Sat.) has little impact on the lower bound for climate sensitivity but, again, increases the upper bound quite considerably (Table 2). The covariance in this case provides a strong constraint on the total outgoing radiation though not on the individual components. Using this analyse we would report a 2.5-97.5% climate sensitivity of 2.6-5.4K with best estimates around 3.3-3.6K and considerable sensitivity to prior assumptions. The sample sensitivity case gives very similar results to the original CERES cases though with an increase of 0.1 K in the overall lower and upper ranges.

Turning now to the 2002 case. Here we only use one year of simulated data to compare with one year of CERES data. Considering the climate sensitivity as a function of RSR and OLR (Fig. 8) we see a very similar plot with higher climate sensitivities at low values of RSR and smallest climate sensitivities at high RSR/OLR. The posterior distributions of climate sensitivity using only this year are similar to the reference CERES case with the same range of 2.7-4.2K. This suggests we could have done our analysis with two year simulations (1 year to spinup and 1 year to compare with observations) rather than the 6 and a half years we actually did. However, we would have still need to have estimates of the climate sensitivity for those configurations.

One other observation is that ERBE and CERES results are quite different from one another. Using the CERES observations and changing covariances changes the upper range of the CDFs but the CDFs are all similar to one below the 60-80% level. The ERBE CDF appears to be characterised by a general shift towards higher sensitivities with differences between it and the CERES distributions apparent at all levels. This leads to differences in the best estimate climate sensitivity giving 3.3 K for the CERES observations and about 4K, though with considerable sensitivity to the prior, for the ERBE observations.

4. Discussion and Conclusions

Having shown that there is a relationship between the two components of outgoing radiation and climate sensitivity we now consider total outgoing radiation. Climate models show on average that:

$$R' = \alpha T' - G \tag{5}$$

Where R' is the change in outgoing radiation, T' the change in surface temperature, G the forcing and α is termed the "climate feedback parameter" (Gregory and Webb 2008). α is related to the climate sensitivity $G_{2\times CO_2}$ (the forcing from doubling CO_2) by $G_{2\times CO_2}/\alpha$. We might expect that with fixed SSTs and, thus, largely constant surface temperatures that increasing α would lead to an increase in total outgoing radiation while decreasing α would lead to a decrease in total outgoing radiation. If this were the case that would provide physical justification for our focus on outgoing radiation in order to observationally constrain climate sensitivity. However, Gregory and Webb (2008); Andrews and Forster (2008); Colman and McAvaney (2011) have shown that CO_2 forcing can also generate, in a model dependant way, rapid changes in tropospheric structure and clouds. The impact of this process could explain why one perturbed physics version of HadSM3 had a low sensitivity (Gregory and Webb 2008). Initially, we neglect this process and assume that the forcing from doubling CO_2 is 3.76 Wm^{-2} (Myhre et al. 1998).

Using this forcing we compute α from the emulated climate sensitivities.

There is an approximately linear, though noisy, relationship between simulated total outgoing radiation and α (Fig. 9) with large outgoing radiation associated, as expected, with large values of α. The slope of the best fit robust line is approximately 30 K i.e. for a 0.1 Wm^{-2} K^{-1} increase in climate feedback strength the outgoing radiation increases by about 3 Wm^{-2} . For climate feedback parameters values greater than about 1.2 Wm^{-2} K^{-1} the regression slope appears to be stronger with a value of about 75 K (Fig. 9). This increase in slope at high climate feedback values cannot be explained by our neglect of rapid clouds response to CO_2 changes. Gregory and Webb (2008) found for one low climate sensitivity configuration of HadSM3 that rapid cloud responses caused the effective forcing to decrease and thus the estimated value of α to reduce. This process, if anything, would steepen the best fit regression slope for the low climate sensitivity cases.

To test if these results arise because of the use of four parameters or because of "tuning" the model to observations we used data from an ensemble of 100 randomly sampled configurations from the 14,001 we used to generate our emulator (see part 1). The 98 cases that did not fail due to numerical problems also show an increase, with considerable scatter, in outgoing radiation as the climate feedback parameter ((Fig. 9). Thus, at least for HadAM3, there appears to be a link between outgoing radiation and climate sensitivity supporting our initial focus on outgoing radiation to constrain climate sensitivity.

Our comparison between model and observation is testing the model fidelity of global-averaged outgoing radiation. With our experimental design it is not possible to test the relative importance of temperature feedbacks and rapid responses to CO_2 and other forcings.

Could we have carried out our analyses more efficiently? We generated about 2500 configurations of HadAM3 and ran each of them for six years. This is computationally expensive and is only possible as HadAM3 is a relatively cheap model. We have already shown that 1 year of data is enough to make reasonable estimates of climate sensitivity. One way to proceed might be to generate the 16 extreme cases and use the three most extreme of these cases to start a series of optimisation cases. To explore this we sub-sampled our data to only include the CERES, ERBE and three targets on the edge of the model-observation consistent region. We then only considered the 2002 data in the analysis. If we had done this we would have concluded that climate sensitivity lay in the range 2.8-4.4K with some sensitivity to prior assumptions. This is not hugely different from our results with 2500 simulations each ran for six years. However, we would still need to compute climate sensitivity for those configurations.

Our focus has been on climate sensitivity for which key processes and parameters have already been identified by Knight et al. (2007). Other impacts of climate change may be less obviously related to present day observations than climate sensitivity appears to be. However, Fowler et al. (2010) found that changes in UK extreme precipitation were strongly controlled by the `entcoef` and `vf1` parameters suggesting our results might also provide some constraints on changes in future extreme precipitation.

We have shown that CERES observations of Reflected Shortwave Radiation (RSR) and Outgoing Longwave Radiation (OLR) provide a significant constraint on the plausible (2.5%-97.5%) range of climate sensitivities for HadAM3. Using the more recent CERES observations we find a range of 2.7-4.2K for HadAM3 climate sensitivity with little sensitivity to a range of prior distributions and a best estimate of 3.4K. Using the older ERBE observations we find greater sensitivity to the prior distribution and a range of 2.8-5.6K with a best estimate around 4K. Amplifying the CERES OLR and RSR errors to make ERBE and CERES observationally consistent leads to high uncertainty in the individual components but, still, small uncertainty in the total outgoing radiation arising from uncertainty in the ocean heating rate and Total Solar Irradiance. This uncertainty estimate gives a best estimate of about 3.4K and a climate sensitivity range of 2.7-5.4K.

Some caveats on our results are necessary. We may be missing or underestimating key uncertainties in model-observational comparison. For example we have assumed that internal climate variability as simulated by default HadAM3 is adequate. The model does not simulate well the average land-surface temperature nor the clear sky outgoing radiation regardless of tuning so if we had included these in our analysis may have reached different conclusions.

Our uncertainty range only includes the effect of atmospheric and land-surface processes and does not take account of oceanographic processes such as changes in ocean circulation nor of changes in sea-ice. However, Brierley et al. (2010) suggest that perturbations in ocean parameters have little impact on future climate change in HadCM3 suggesting our neglect of them is not critical. Thus, our results suggest that climate sensitivity, for the HadAM3 model, is unlikely (2.5%) to be greater than 5.6K. This uncertainty could be narrowed given focused work by the satellite community to resolve differences between the CERES and ERBE measurements.

Our results also have implications for the recent UK Climate Projections(Murphy et al. 2010) analysis which is based on a set of 11 perturbed physics regional model simulations of HadAM3 driven by perturbed physics simulations of the HadCM3 Atmosphere/Ocean General Circulation Model(Collins et al. 2011). All these configurations of HadCM3 require flux correction and some have climate sensitivities larger than our results suggest is plausible. If

our results are correct then the impact of climate change may be less severe than some of those simulations suggest.

Other groups have found quite different results for the range of plausible climate sensitivities. Shiogama et al. (2012) report a climate sensitivity range of 2.2-3.2K for the MIROC5 model. However, unlike our range, their range is not based on a measure of model-observational difference but instead that the atmospheric model should have a small net TOA imbalance when driven with SSTs taken from a pre-industrial control simulation. This experimental design is likely to underestimate the range of climate sensitivities as any coupled model configuration ran to equilibrium would have a small net top of atmospheric imbalance. So it is possible that other configurations, with a broader range of climate sensitivities, when ran with different pre-industrial SSTs would have also been in radiation balance.

Sanderson (2011), using the CAMcube model, also found a narrow range of climate sensitivities (2.2-3.2K) by perturbing four parameters across their plausible values. No observational constraints were applied which, presumably, would reduce the range still further.

Our study used HadAM3 and varied four parameters which previous work suggested was important in the models climate sensitivity. Thus, our results are conditional on both the model and the parameters varied. Our uncertainty estimates, based as they are on HadAM3, largely do not include structural uncertainty for which using additional models is one way forward. As the work described above suggests different models are likely to produce different ranges of climate sensitivities. One way forward is to generate, for each model, a range of perturbed models consistent with observations. If this can be done efficiently then this would allow a better understanding of the range of possible future climates in response to emissions of greenhouse gases.

Acknowledgments.

SFBT was supported by the National Centre for Earth Observations (NERC grant NE/F001436/1). Support for MJM and computer time on the Edinburgh Computing and Data Facility was provided by the Centre for Earth Dynamics which is part of the Scottish Alliance for GeoScience, Environment and Society. CC, SFBT and MJM also acknowledge support from "Bridging the Gaps" and "Maximaths" funding. DJR was supported by a NERC PhD studentship with a CASE award from CEH Wallingford. We are grateful to volunteers who participated in the climate*prediction*.net experiment, donating spare computing cycles to integrate the HadSM3 model versions, from which we were able to calculate the climate sensitivity for each model parameter configuration. We thank Joyce Penner for providing results from simulations used to estimate the uncertainty in natural aerosols effect on RSR.

REFERENCES

Andrews, T. and P. M. Forster, 2008: CO2 forcing induces semi-direct effects with consequences for climate feedback interpretations. *Geophys. Res. Lett.*, **35 (4)**, doi:10.1029/2007GL032273.

Aurenhammer, F., 1991: Voronoi Diagrams - A Survey Of A Fundamental Geometric Data Structure. *Computing Surveys*, **23 (3)**, 345–405, doi:ISI:A1991GM03300003.

Breiman, L., 2001: Random Forests. *Machine Learning*, **45 (1)**, 5–32.

Brierley, C. M., M. Collins, and A. J. Thorpe, 2010: The impact of perturbations to ocean-model parameters on climate and climate change in a coupled model. *Clim. Dyn.*, **34 (2-3)**, 325–343, doi:10.1007/s00382-008-0486-3.

Carslaw, K. S., O. Boucher, D. V. Spracklen, G. W. Mann, J. G. L. Rae, S. Woodward, and M. Kulmala, 2010: A review of natural aerosol interactions and feedbacks within the Earth system. *Atmos. Chem. Phys.*, **10 (4)**, 1701–1737.

Collins, M., B. Booth, B. Bhaskaran, G. Harris, J. Murphy, D. Sexton, and M. Webb, 2011: Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. *Clim Dyn.*, **36 (9)**, 1737–1766.

Colman, R. A. and B. J. McAvaney, 2011: On tropospheric adjustment to forcing and climate feedbacks. *Clim. Dyn.*, **36 (9-10)**, 1649–1658, doi:10.1007/s00382-011-1067-4.

Fasullo, J. T. and K. E. Trenberth, 2008: The annual cycle of the energy budget. Part I: Global mean and land-ocean exchanges. *J. Clim.*, **21 (10)**, 2297–2312, doi:10.1175/2007JCLI1935.1.

Forster, P., et al., 2007: Changes in atmospheric constituents and in radiative forcing. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group 1 to the Fourth Assessment report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, 129–234.

Fowler, H. J., D. Cooley, S. R. Sain, and M. Thurston, 2010: Detecting change in UK extreme precipitation using results from the climateprediction.net BBC climate change experiment. *Extremes*, **13 (2)**, 241–267, doi:10.1007/s10687-010-0101-y.

Gregory, J. M. and M. J. Webb, 2008: Tropospheric adjustment induces a cloud component in CO_2 forcing. *J. Climate*, **21**, 58–71.

Hegerl, G. C., T. J. Crowley, W. T. Hyde, and D. J. Frame, 2006: Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature*, **440**, 1029–1032.

Huber, M., I. Mahlstein, M. Wild, J. Fasullo, and R. Knutti, 2011: Constraints on Climate Sensitivity from Radiation Patterns in Climate Models. *J. Clim.*, **24 (4)**, 1034–1052, doi:10.1175/2010JCLI3403.1.

Huybers, P., 2010: Compensation between model feedbacks and curtailment of climate sensitivity. *J. Clim.*, **23**, 3009–3018, doi:10.1175/2010JCLI3380.1.

Jackson, C. S., M. K. Sen, G. Huerta, Y. Deng, and K. P. Bowman, 2008: Error Reduction and Convergence in Climate Prediction. *J. Clim.*, **21 (24)**, 6698–6709, doi:10.1175/2008JCLI2112.1.

Järvinen, H., P. Räisänen, M. Laine, J. Tamminen, A. Ilin, E. Oja, A. Solonen, and H. Haario, 2010: Estimation of ECHAM5 climate model closure parameters with adaptive MCMC. *Atmos. Chem. Phys*, **10**, 9993–10 002.

Kettleborough, J. A., B. B. B. Booth, P. A. Stott, and M. R. Allen, 2007: Estimates of uncertainty in predictions of global mean surface temperature. *J. Clim.*, **20 (5)**, 843855.

Knight, C. G., et al., 2007: Association of parameter, software, and hardware variation with large-scale behavior across 57,000 climate models. *Proc. Natl. Acad. Sci. U. S. A.*, **104 (30)**, 12 259–12 264, doi:10.1073/pnas.0608144104.

Knutti, R. and G. Hegerl, 2008: The equilibrium sensitivity of the Earth's temperature to radiation changes. *Nature Geoscience*, **1 (11)**, 735–743.

Kopp, G. and J. L. Lean, 2011: A new, lower value of total solar irradiance:evidence and climate significance. *Geophys. Res. Lett.*, L01706, doi:10.1029/2010GL045777.

Lemoine, D., 2010: Climate sensitivity distributions dependence on the possibility that models share biases. *Journal of Climate*, **23 (16)**, 4395–4415, doi:10.1175/2010JCLI3503.1.

Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, 2009: Toward optimal closure of the earth's top-of-atmosphere radiation budget. *J. Clim.*, **22**, 748–765, doi:10.1175/2008JCLI2637.1.

Lyman, J. M., S. A. Good, V. V. Gouretski, M. Ishii, G. C. Johnson, M. D. Palmer, D. M. Smith, and J. K. Willis, 2010: Robust warming of the global upper ocean. *Nature*, **465 (7296)**, 334–337, doi:10.1038/nature09043.

Meehl, G. A., et al., 2007: Global climate projections. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Eds., Cambridge University Press.

Meinshausen, N., 2006: Quantile regression forests. *J. M. L. R.*, **7**, 983–999.

Murphy, J., B. Booth, M. Collins, G. Harris, D. Sexton, and M. Webb, 2007: A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philos. Trans. R. Soc. London*, **365**, 2133–2133.

Murphy, J., et al., 2010: UK Climate Projections science report: Climate change projections. Tech. rep., UK Climate Projections. URL `http://ukclimateprojections.defra.gov.uk/media.jsp?mediaid=87893&filetype=pdf`, accessed July 21st, 2012.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772.

Myhre, G., E. J. Highwood, K. P. Shine, and F. Stordal, 1998: New estimates of radiative forcing due to well mixed greenhouse gases. *Geophys. Res. Lett.*, **25 (14)**, 2715–2718.

Olson, R., R. Sriver, M. Goes, N. Urban, H. Matthews, M. Haran, and K. Keller, 2012: A climate sensitivity estimate using Bayesian fusion of instrumental observations and an Earth System model. *Journal of Geophysical Research*, **117 (D4)**, D04 103, doi:10.1029/2011JD016620.

Penner, J. E., et al., 2006: Model intercomparison of indirect aerosol effects. *Atmos. Chem. Phys.*, **6**, 3391–3405.

Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton, 2000: The impact of new physical parametrizations in the Hadley Centre climate model – HadAM3. *Clim. Dyn.*, **16**, 123–146.

Randall, D. A., et al., 2007: Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, chap. 8, 589–662.

Rayner, N., P. Brohan, D. E. Parker, C. K. Folland, J. Kennedy, M. Vanicek, T. Ansell, and S. F. B. Tett,

2006: Improved analyses of changes and uncertainties in sea surface temperature measured *in situ* since the mid-nineteenth century. *J. Climate*, **19**, 446–469.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Q. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15 (13)**, 1609–1625.

Rougier, J., 2007: Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change*, **81 (3)**, 247–264.

Rougier, J., D. M. H. Sexton, J. M. Murphy, and D. Stainforth, 2009: Analyzing the Climate Sensitivity of the HadSM3 Climate Model Using Ensembles from Different but Related Experiments. *J. Clim.*, **22 (13)**, 3540–3557, doi:10.1175/2008JCLI2533.1.

Roulston, M. and L. Smith, 2003: Combining dynamical and statistical ensembles. *Tellus*, **55 (1)**, 16–30.

Rowlands, D., et al., 2012: Broad range of 2050 warming from an observationally constrained large climate model ensemble. *Nature Geoscience*, **5 (4)**, 256–260.

Sanderson, B. M., 2011: A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. *J. Clim.*, **34**, 1362–1377, doi:10.1175/2010JCLI3498.1.

Sanderson, B. M., et al., 2008: Constraints on model response to greenhouse gas forcing and the role of subgrid-scale processes. *J. Clim.*, **21 (11)**, 2384–2400, doi:10.1175/2008JCLI1869.1.

Sanso, B. and C. Forest, 2009: Statistical calibration of climate system properties. *J. Roy. Statistical Society Series (Applied Statistics)*, **58 (Part 4)**, 485–503.

Sexton, D. M. H. and J. M. Murphy, 2011: Multivariate probabilistic projections using imperfect climate models. Part II: robustness of methodological choices and consequences for climate sensitivity. *Clim. Dyn.*, doi:10.1007/s00382-011-1209-8.

Shiogama, H., et al., 2012: Perturbed physics ensemble using the MIROC5 coupled atmosphere–ocean gcm without flux corrections: experimental design and results. *Climate Dynamics*, **39**, 1–16, doi:10.1007/s00382-012-1441-x.

Solomon, S., et al., 2007: *Climate Change 2007: The physical Science Basis. Contribution of Working Group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, chap. IPCC, 2007: Technical Summary. Cambridge University Press.

Stainforth, D. A., et al., 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406.

Tett, S. F. B., et al., 2002: Estimation of natural and anthropogenic contributions to 20^{th} century temperature change. *J. Geophys. Res.*, **107**, doi:10.1029/2000JD000028.

Tett, S. F. B., et al., 2007: The impact of natural and anthropogenic forcings on climate and hydrology. *Clim. Dyn.*, **28 (1)**, 3–34, doi:10.1007/s00382-006-0165-1.

Webb, M. J., et al., 2006: On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. *Clim. Dyn.*, **27 (1)**, 17–38, doi:10.1007/s00382-006-0111-2.

Williams, K. D., C. A. Senior, and J. F. B. Mitchell, 2001: Transient climate change in the Hadley Centre models: The role of physical processes. *J. Climate*, **14 (12)**, 2659–2674.

Willis, J., D. Roemmich, and B. Cornuelle, 2004: Interannual variability in upper ocean heat content, temperature, and thermosteric expansion on global scales. *J. Geophys. Res.-Oceans*, **109 (C12)**, doi:10.1029/2003JC002260.

Willson, R. and H. Hudson, 1991: The sun's luminosity over a complete solar cycle. *Nature*, **351 (6321)**, 42–44.

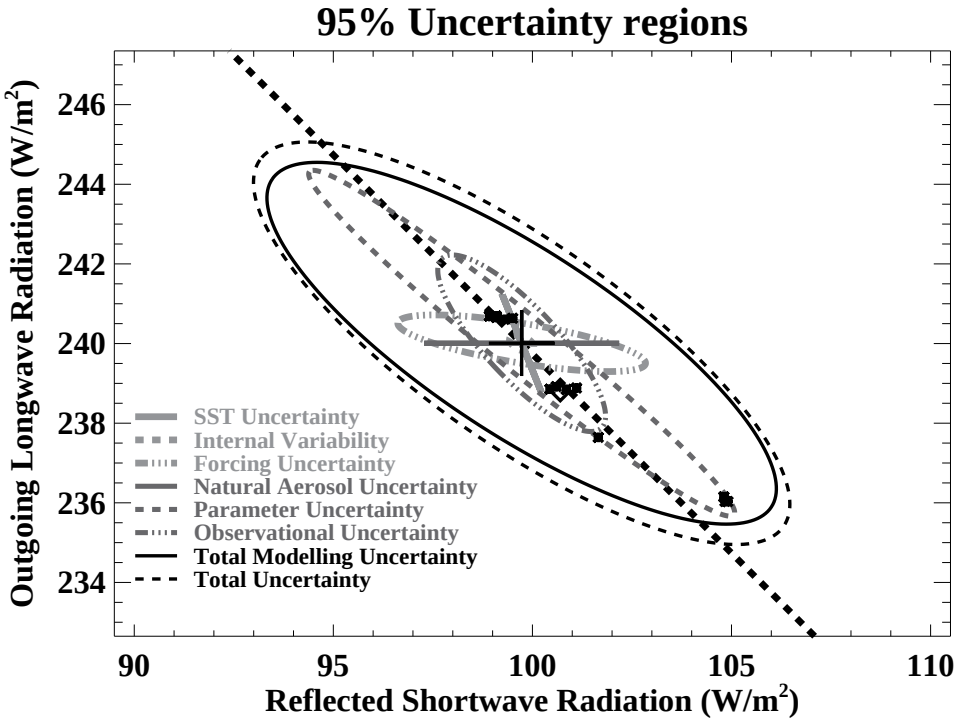

FIG. 1. 95% uncertainty regions for different sources of uncertainty (see key). Also shown are parameter combinations with default values for `entcoef`, `vf1`, `ct` & `rhcrit` and a climate sensitivity in the range 3.2-3.4K (stars) and the default configuration of HadAM3 (black diamond).

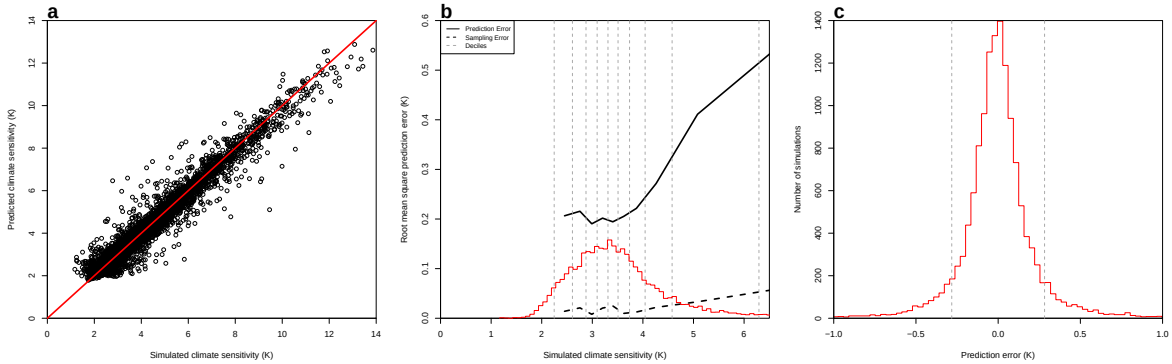

FIG. 2. Validation of the randomForest statistical model used to predict climate sensitivities. a) Results from a 10-fold cross validation over the 14,001 member training set from climate*prediction*.net. Shown are the out-of-sample predicted climate sensitivities against the simulated values from HadSM3. Red line shows the theoretical 1:1 relationship, b) Root mean square prediction error as a function of simulated climate sensitivity (solid black line). Also shown is the distribution of simulated climate sensitivities (red histogram), and root mean error variance due to sampling uncertainty (dashed black line). Dashed grey lines indicate the deciles of the distribution defining the bins. c) Histogram of prediction error from the out-of-sample predictions. The root mean square error of ≈ 0.3K is shown by dotted grey lines

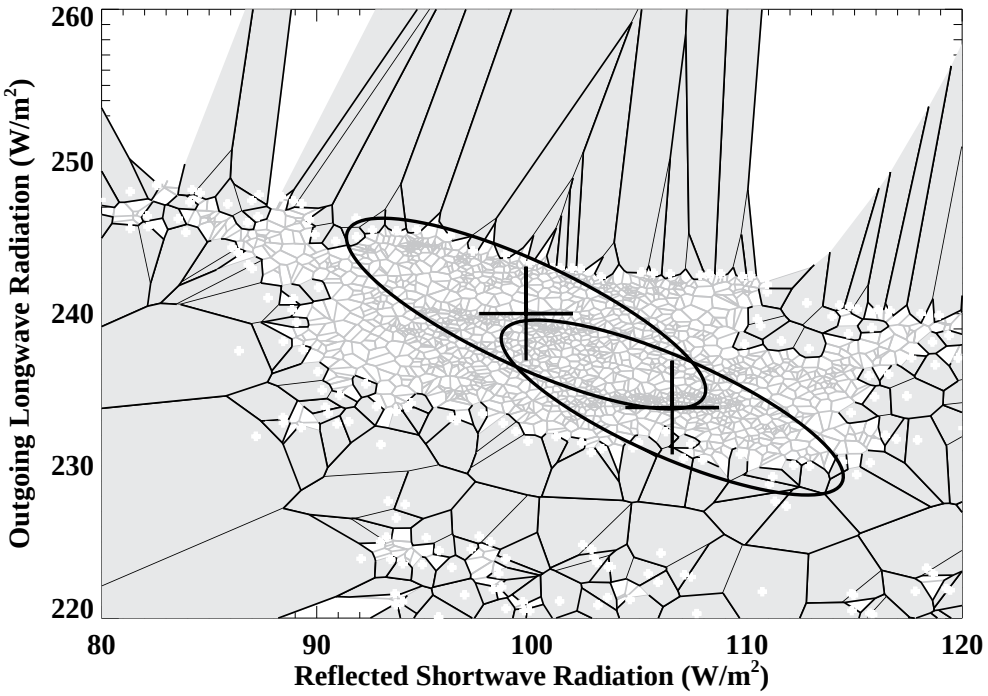

FIG. 3. Voronoi Polygons for modified physics simulations with only those polygons in the range 80-120 Wm^{-2} (RSR) and 220-260 Wm^{-2} (OLR) are shown. Light gray filled polygons are those that have an area greater than π. These polygons have their areas capped at value of π for subsequent computations. Also shown are the 99% uncertainty estimates (black ellipses) centred on CERES, at about (100, 240) Wm^{-2} , and ERBE, at about (107,234) Wm^{-2} , observations (black crosses).

TABLE 1. Change in outgoing radiation (Wm^{-2}) between experiment and ensemble average of standard HadAM3 configuration. Forcing column shows anthropogenic forcing from Tett et al. (2007) with RSR being the SW forcing and OLR being the LW forcing. SW forcing largely arises from aerosols while LW forcing largely arises from well mixed greenhouse gases. However, ozone and land-use changes affect both the SW and LW forcing.

	Experiment			
	NOSUL	**NATURAL**	**NATGHG**	**Anthro. Forcing**
RSR	-1.18	-0.83	-1.55	-0.93
OLR	0.19	2.64	0.58	2.11

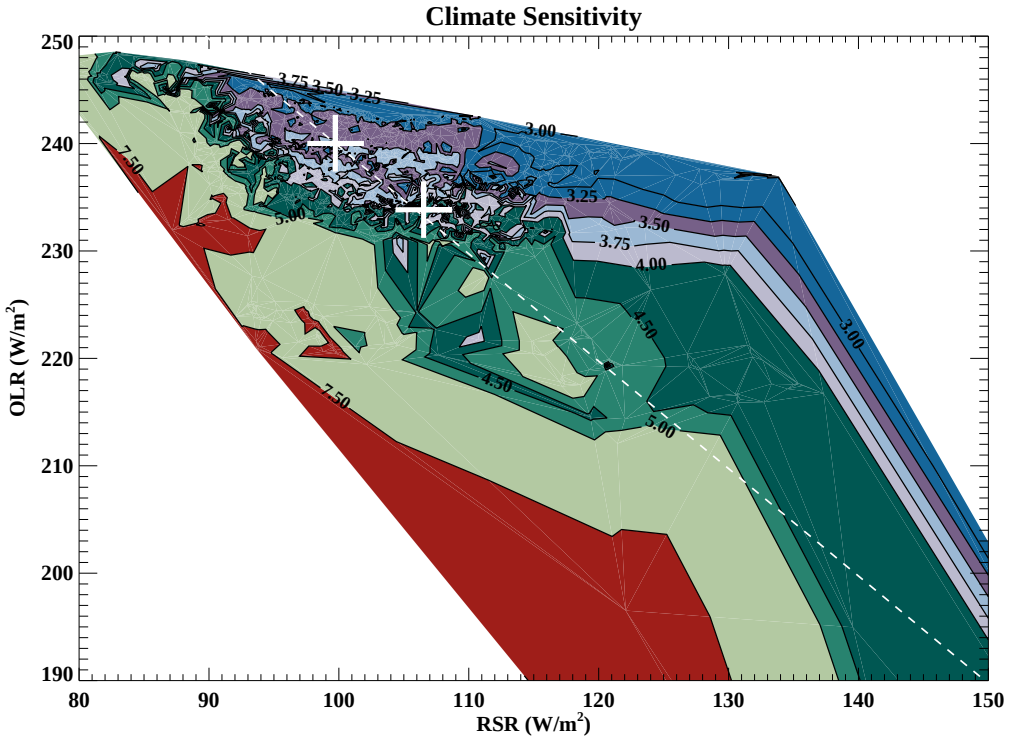

FIG. 4. Climate Sensitivity (K & colours) as a function of RSR (x-axis) and OLR (y-axis). The two white crosses show the CERES (upper) and ERBE (lower) values while dashed diagonal line show total outgoing radiation agreeing with near radiative balance. Contour levels are at 2, 2.5, 3,3.25,3.5,3.75, 4, 4.5 and 7K

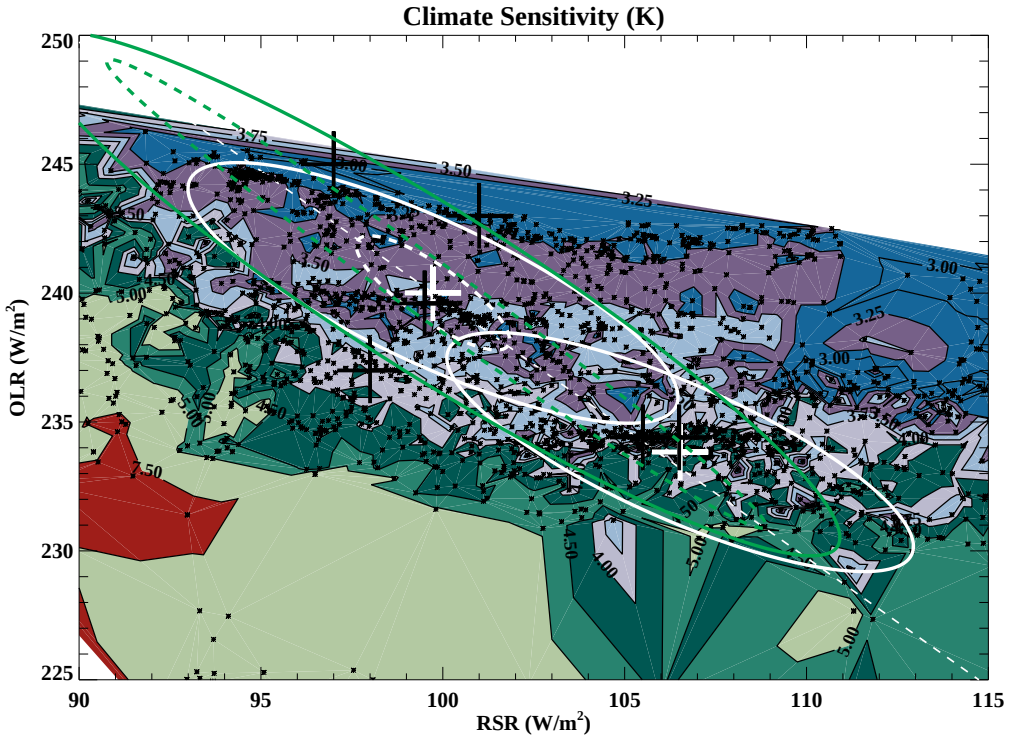

FIG. 5. Climate Sensitivity (K & colours) as a function of RSR (x-axis) and OLR (y-axis). Details as Fig. 4 but focusing on the region 90–115 Wm^{-2} (RSR) and 225–250 Wm^{-2} (OLR). Also shown is the region where simulated OLR and RSR is consistent with CERES or ERBE observations at the 95% level (white ellipse). The thick dashed ellipse centred on the CERES observations shows the 95% consistency region when only observational uncertainty is considered. The green dashed ellipse is the observational uncertainty region when the satellite errors are scaled by a factor of 20 (see main text). The solid green line is the total uncertainty in this case. The six black crosses show the optimisation targets.

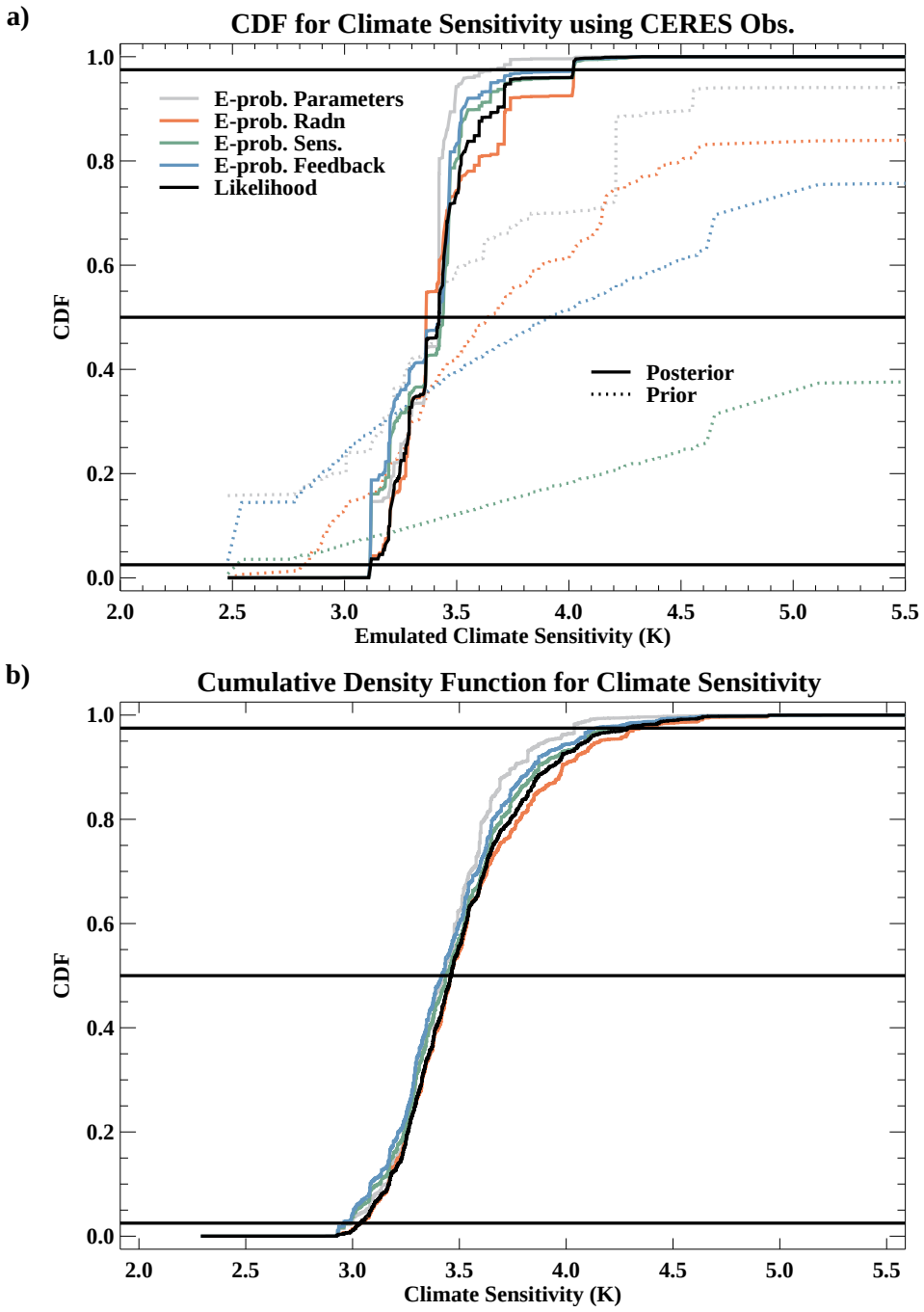

FIG. 6. Cumulative distribution functions for climate sensitivity (S). a) prior (dotted) and posterior (solid) cumulative distribution functions for five different prior distributions which are Equi-probable in: parameters (gray), radiation (red), climate sensitivity (green), feedback strength (blue). Solid black line shows the cumulative likelihood distribution. Solid horizontal lines show the 2.5, 50, and 97.5 % values. b) as a) but for posterior distributions when uncertainty in emulated estimate of S is included. Other details as a)

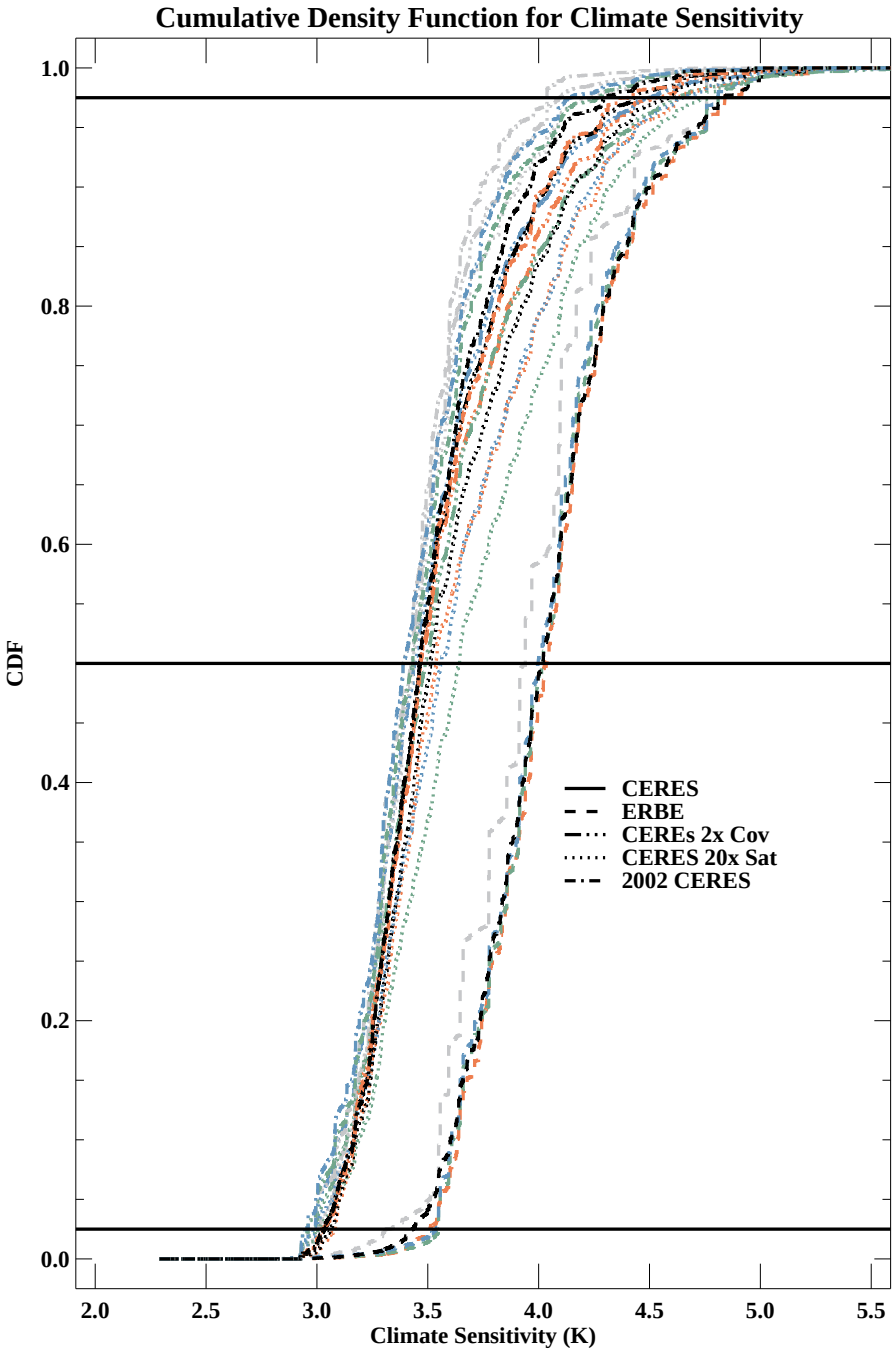

FIG. 7. Cumulative distribution functions for climate sensitivity (S) for a variety of sensitivity studies (see key). Other details as Fig. 6(b).

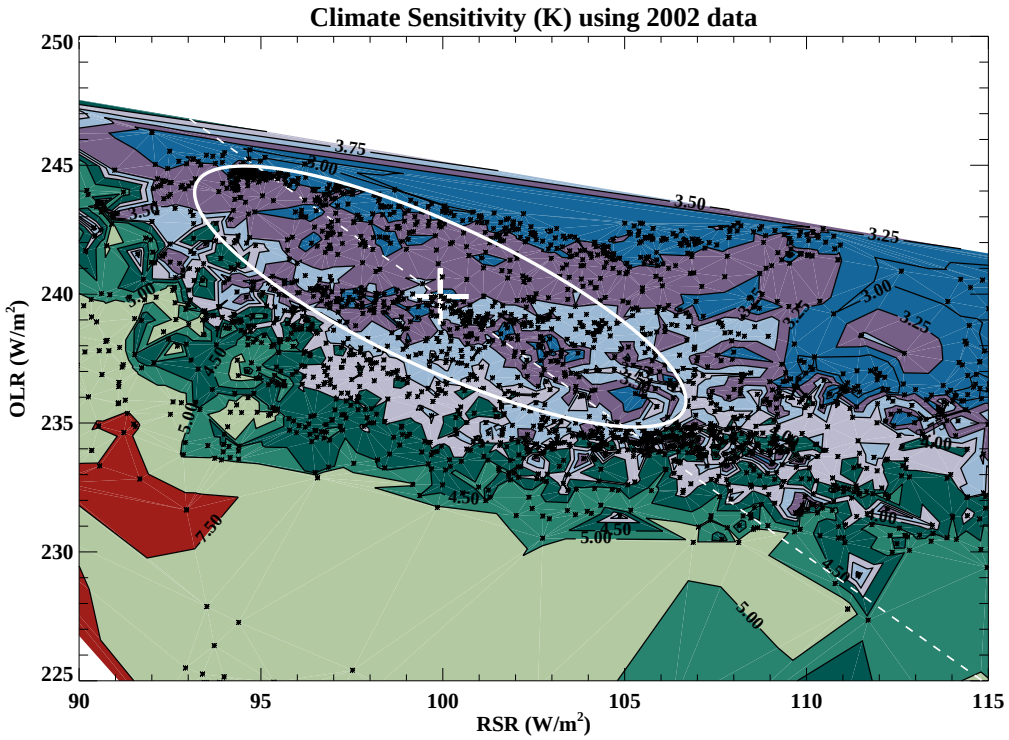

FIG. 8. As Fig. 5 but for 2002 data and uncertainties.

TABLE 2. Climate Sensitivity Summary. Shown for each named observation, uncertainty estimate and prior (see main text) are the best estimate, 2.5% and 97.5% estimates of climate sensitivity (K) for HadAM3. Values are taken from the posterior distributions.

Name	Prior	Best Est.	2.5%	97.5%	Name	Best Est.	2.5%	97.5%
CERES	Uniform	3.4	2.8	4.2	ERBE	4.0	3.0	5.3
	Parameters	3.4	2.8	4.1		3.7	2.8	4.8
	Radiation	3.4	2.7	4.1		4.3	3.0	5.6
	S	3.4	2.8	4.2		4.3	3.0	5.6
	1/S	3.4	2.8	4.2		4.1	2.9	5.5
CERES 2×	Uniform	3.5	2.8	4.4	CERES 20× Sat.	3.5	2.8	4.6
	Parameters	3.4	2.7	4.2		3.3	2.7	4.1
	Radiation	3.3	2.6	4.4		3.4	2.6	5.0
	S	3.5	2.7	5.0		3.6	2.7	5.4
	1/S	3.4	2.7	4.7		3.4	2.7	5.0
2002 CERES	Uniform	3.4	2.8	4.2	CERES Sample	3.4	2.8	4.1
	Parameters	3.4	2.7	4.1		3.4	2.9	4.1
	Radiation	3.4	2.7	4.1		3.4	2.8	4.2
	S	3.4	2.8	4.3		3.4	2.8	4.2
	1/S	3.4	2.8	4.2		3.3	2.7	4.2

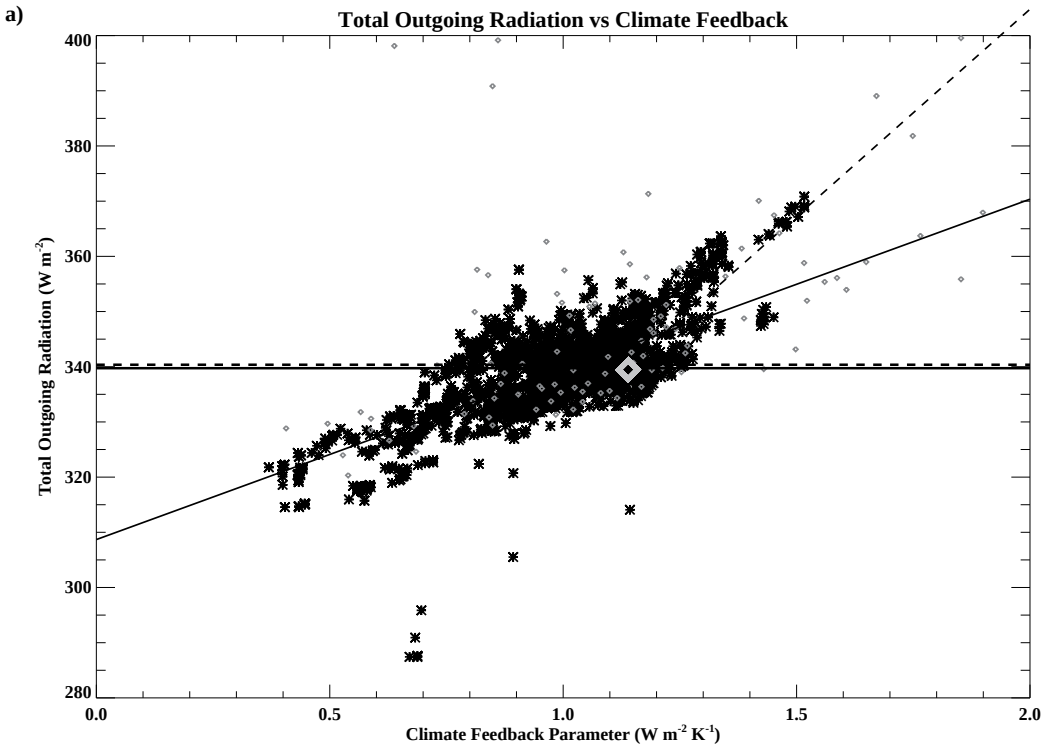

FIG. 9. Total Outgoing Radiation (y-axis) vs climate feedback (x-axis). Asterisks show results for individual simulations. Dark gray diamonds show results from the random sample ensemble. Climate feedback is computed using a forcing of 3.76 Wm^{-2} for doubling CO_2. Results from default parameter simulation of HadAM3 are shown with a gray diamond assuming a default climate sensitivity of 3.3K. The solid horizontal line shows the total outgoing radiation from Loeb et al. (2009) while the dashed horizontal line shows the same from Fasullo and Trenberth (2008).